

MBC 638

DATA ANALYSIS AND DECISION MAKING

Anna Chernobai, Department of Finance

**CONFIDENCE INTERVAL for the difference in
population means**
Chapter 8

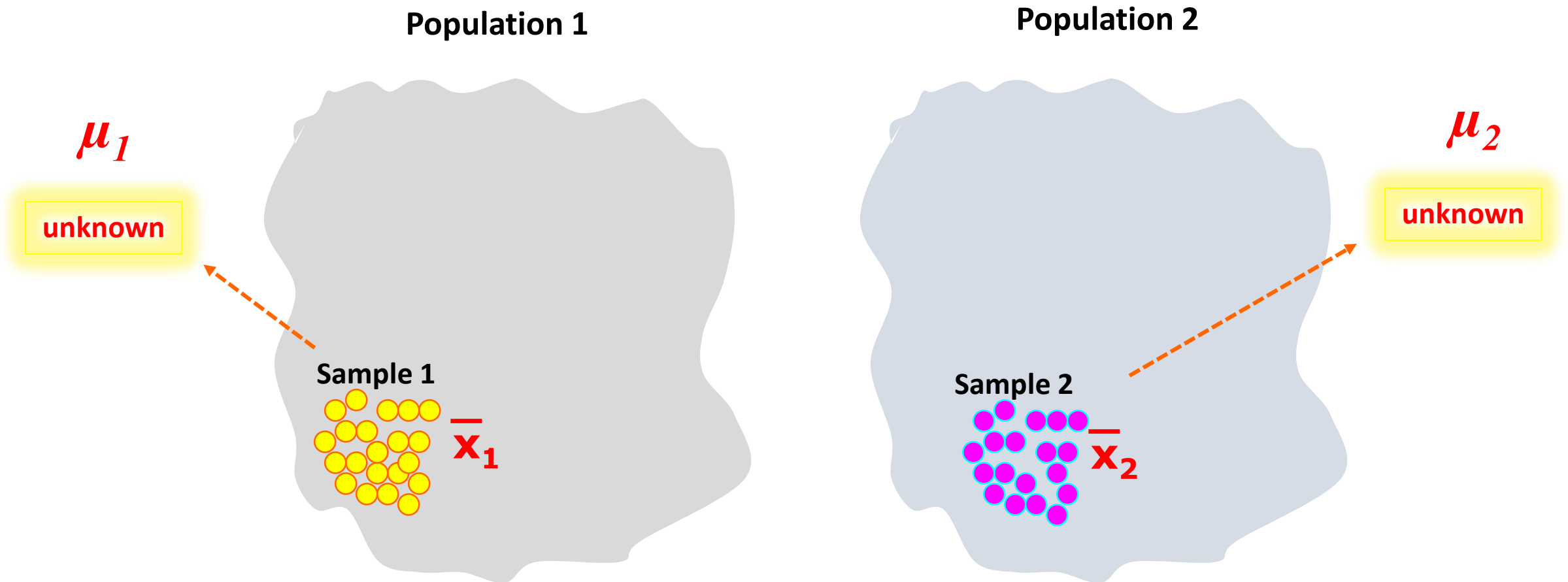
Today: Confidence interval for $\mu_1 - \mu_2$

- Independent samples
- Dependent samples

Confidence interval for $\mu_1 - \mu_2$

- ... will help you answer questions such as:
 - Do Toyota cars cost higher in California than New York?
 - Have the NY city apartment prices risen significantly since last year?
 - Do TV advertisements help sell your product better?
 - Is there a gender difference in the preferences regarding a product?
 - Does having an MBA degree help increase salary? Job satisfaction?
- To answer such questions, we would need data from 2 samples taken from 2 different populations.

- We use information from 2 observed samples to conclude how μ_1 and μ_2 compare.
- Assumption: The two samples are **independent**. (We will relax this assumption later.)



Recall:

- \bar{X} is a **point estimate** of μ
- \bar{X} is a **random variable**
- μ is the **parameter** that we want to estimate

$(1 - \alpha)\%$ C.I. for μ is:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Standard deviation of \bar{X}

- \bar{X} is a **point estimate** of μ
- \bar{X} is a **random variable**
- μ is the **parameter** that we want to estimate

$(1 - \alpha)\%$ C.I. for μ is:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Standard deviation of \bar{X}

- \bar{X} is a **point estimate** of μ
- \bar{X} is a **random variable**
- μ is the **parameter** that we want to estimate

$(1 - \alpha)\%$ C.I. for μ is:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Standard deviation of \bar{X}

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

By the same logic, to construct a confidence interval for $\mu_1 - \mu_2$:

■ $\bar{x}_1 - \bar{x}_2$ is the **point estimate**

■ $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ is the **standard deviation** of $\bar{X}_1 - \bar{X}_2$ (the two samples are independent)

Note:

Whenever sample sizes are small, we always assume **bell-shaped** populations for X1 and X2

- Case 1: σ_1, σ_2 are known: Use Z

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- Case 2: σ_1, σ_2 are unknown, $n_1, n_2 \geq 30$: Use **t** (exact), **d.f.= n_1+n_2-2** or Z (approximation)

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

or

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- Case 3: σ_1, σ_2 are unknown, $n_1, n_2 < 30$: Use **t** , **d.f.= n_1+n_2-2**

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Examples

Example

Periodically, Merrill Lynch customers are asked to evaluate Merrill Lynch financial consultants and services (*2000 Merrill Lynch Client Satisfaction Survey*) . Higher ratings on the client satisfaction survey indicate better service, with 7 the maximum service rating.

Independent samples of service ratings for 2 financial consultants—Ms. Chen and Mr. Wang are summarized here. Ms. Chen has 10 years of experience, whereas Mr. Wang has 1 year of experience.

	Ms. Chen	Mr. Wang
# reviews collected	56	40
Mean	6.82	6.25
Standard deviation	0.64	0.75

At 95% confidence level, determine whether consultants with more experience provide better service.

Confidence interval for $\mu_1 - \mu_2$ in Excel

Example

SAT Verbal.xlsx

Today: Confidence interval for $\mu_1 - \mu_2$

■ Independent samples



■ Dependent samples

What if the samples are **not independent**?

Paired samples

Example

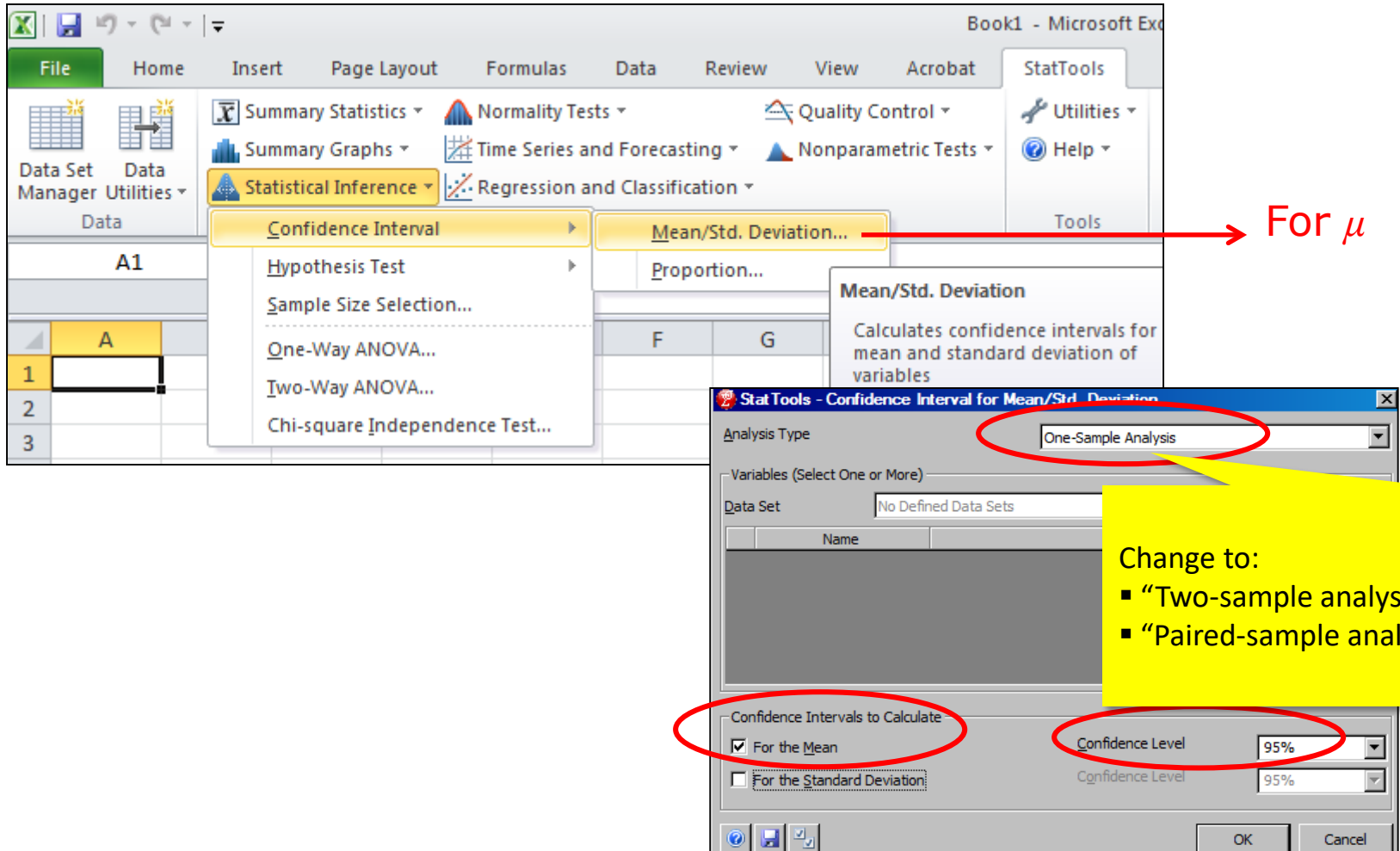
Sales Presentation Ratings.xlsx

Example

Task Times.xlsx

Excel

StatTools → Statistical Inference → Confidence Interval → Mean



The image shows the Excel StatTools ribbon with the 'Statistical Inference' dropdown menu open. The 'Confidence Interval' option is selected, and the 'Mean/Std. Deviation...' dialog box is displayed. A red arrow points from the 'Mean/Std. Deviation...' option in the menu to the text 'For μ '. The dialog box has 'One-Sample Analysis' selected in the 'Analysis Type' dropdown. A yellow callout box points to this dropdown with the text 'Change to:'. Below the callout, a list of options is provided: 'Two-sample analysis' for independent samples and 'Paired-sample analysis' for paired (matched) samples. In the dialog box, the 'Confidence Intervals to Calculate' section has 'For the Mean' checked and circled in red. The 'Confidence Level' is set to 95%.

For μ

StatTools - Confidence Interval for Mean/Std. Deviation

Analysis Type: One-Sample Analysis

Variables (Select One or More):

Data Set: No Defined Data Sets

Confidence Intervals to Calculate:

- ☒ For the Mean
- ☐ For the Standard Deviation

Confidence Level: 95%

Confidence Level: 95%

Change to:

- "Two-sample analysis" for independent samples
- "Paired-sample analysis" for paired (matched) samples

Today: Confidence interval for $\mu_1 - \mu_2$

■ Independent samples



■ Dependent (paired) samples

